

# Generalized Stacked Sequential Learning

Eloi Puertas Prats

*Matemàtica Aplicada i Anàlisi, Universitat de Barcelona, Gran Via 585, Barcelona, Spain*

*Advisor/s: Dr. Oriol Pujol Vila, Dr. Sergio Escalera*

*Date and location of PhD thesis defense: 14 November 2014, Universitat de Barcelona*

Received 23rd March 2015; accepted 14th May 2015

---

## 1 Abstract

In many supervised learning problems, it is assumed that data is independent and identically distributed. This assumption does not hold true in many real cases, where a neighboring pair of examples and their labels exhibit some kind of relationship. Sequential learning algorithms take benefit of these relationships in order to improve generalization. In the literature, there are different approaches that try to capture and exploit this correlation by means of different methodologies. In this thesis we focus on meta-learning strategies and, in particular, the stacked sequential learning (SSL) framework.

The main contribution of this thesis is to generalize the SSL highlighting the key role of how to model the neighborhood interactions. We propose an effective and efficient way of capturing and exploiting sequential correlations that take into account long-range interactions. We tested our method on several tasks: text line classification, image pixel classification, multi-class classification problems and human pose segmentation. Results on these tasks clearly show that our approach outperforms the standard stacked sequential learning as well as off-the-shelf graphical models such conditional random fields.

## 2 Generalized Stacked Sequential Learning [2]

Sequential learning assumes that samples are not independently drawn from a joint distribution of the data samples  $\mathbf{X}$  and their labels  $Y$ . Therefore, here the training data is considered as a sequence of pairs: example and its label  $(x, y)$ , such that neighboring examples exhibit some kind of relationship.

Cohen and Carvalho [1] developed an approach of sequential learning based on the stacking framework called Stacked Sequential Learning (SSL). This scheme is a two layers classifier where, firstly, a base classifier  $H_1(x)$  is trained and tested with the original data  $\mathbf{X}$ . Then, an extended data set is created which joins the original training data features  $\mathbf{X}$  with the predicted labels  $Y'$  produced by the base classifier considering a fixed-size window around the example. Finally, second classifier  $H_2(x)$  is trained with this new feature set. The final result is a set of predictions  $\hat{Y}$ . Our framework for generalizing the stacked sequential learning includes a new block, called  $J$ , in the pipeline of the basic SSL. Figure 1 shows the Generalized Stacked Sequential Learning process.

---

Correspondence to: <epuertas@ub.edu>

Recommended for acceptance by Jorge Bernal

DOI <http://dx.doi.org/10.5565/rev/elcvia.737>

ELCVIA ISSN:1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

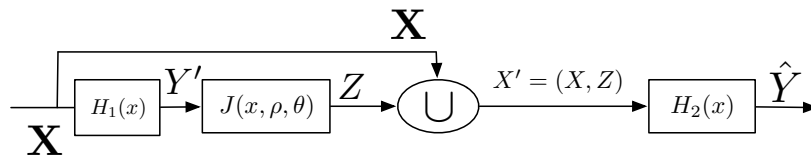


Figure 1: Block diagram for the generalized stacked sequential learning.

Input	Ground truth	$h_1(x)$	SSL 7×7	CRF	MR-SSL

Figure 2: Input, ground truth and results of different methods on horse images.

As before, a classifier  $H_1(x)$  is trained with the input data set  $\mathbf{X} \in (\mathbf{x}, \mathbf{y})$  and the predicted labels  $Y'$  are obtained. Next, the new block  $J$  defines the policy for creating the neighborhood model of the predicted labels, where  $z = J(y', \rho, \theta) : \mathcal{R} \rightarrow \mathcal{R}^w$  is a function that captures the data interaction with a model parameterized by  $\theta$  in a neighborhood  $\rho$ . The result of this function is a  $w$ -dimensional value, where  $w$  is the number of elements in the support lattice of the neighborhood  $\rho$ . Then, the output  $z = J(y', \rho, \theta)$  is joined with the original training data creating the extended training set  $\mathbf{X}' \in (\mathbf{x}', \mathbf{z})$ . This new set is used to train the second classifier  $H_2(\mathbf{x}')$  with the goal of producing the final prediction  $\hat{Y}$ . Observe that the system will be able to deal with neighboring relations depending on how well  $J(y', \rho, \theta)$  characterize them. Figure 2 shows that defining a neighboring relationship function based on multi-resolution decomposition, our framework beats other frameworks like Stacked Sequential Learning (SSL) and Conditional Random Fields (CRF).

## References

- [1] W.W Cohen, V.R. de Carvalho "Stacked sequential learning", *Proceedings of IJCAI 2005*, 671–676, 2005.
- [2] E. Puertas, S. Escalera, O. Pujol "Generalized multi-scale stacked sequential learning for multi-class classification", *Pattern Analysis and Applications*, 1–15, 2013.